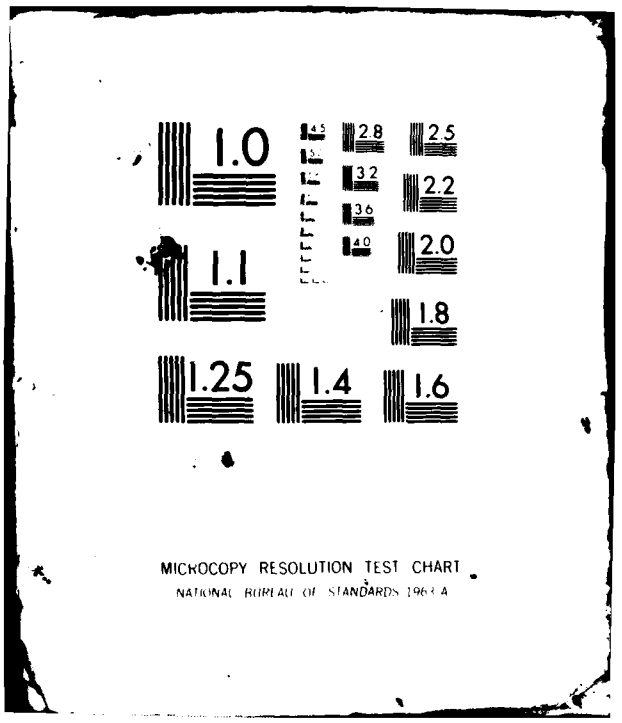
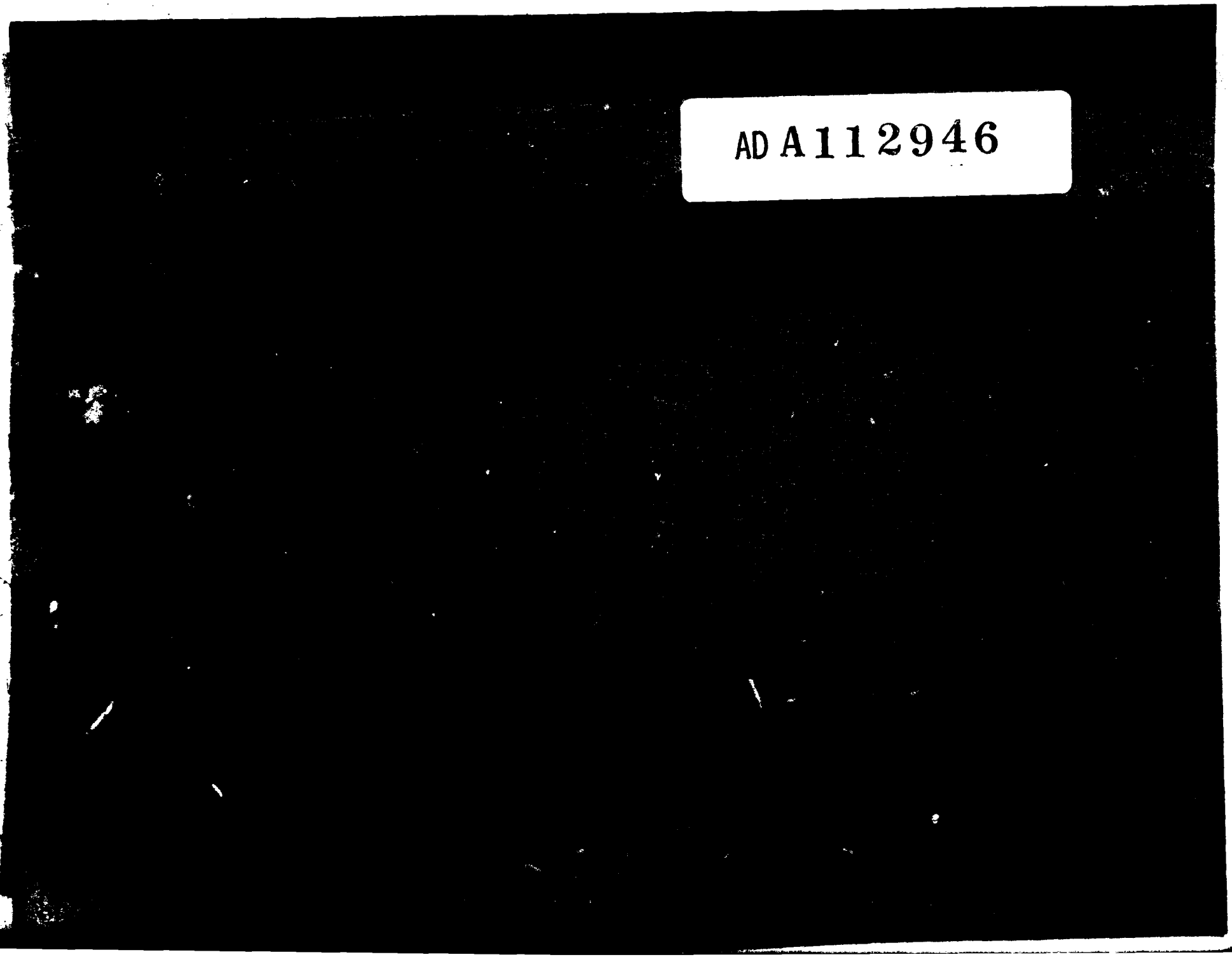
AD-A112 946 CLEMSON UNIV SC DEPT OF MATHEMATICAL SCIENCES F/G 12/1
A BAYES RULE FOR SELECTING THE LARGEST COMPONENT.(U)
JUL 81 K ALAM N00014-75-C-0451
UNCLASSIFIED N127 NL

END
DATE FILMED
5 82
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A



(12)

A BAYES RULE FOR SELECTING THE LARGEST COMPONENT

By

Khursheed Alam*

N127

Technical Report #365

July 16, 1981

Clemson University

*This work was supported by the U. S. Office of Naval Research under Contract N00014-75-C-0451.

A BAYES RULE FOR SELECTING THE LARGEST COMPONENT

Khursheed Alam*

Clemson University

ABSTRACT

Let $\underline{X} = (X_1,\ldots,X_k)$ be a random vector whose distribution depends on an unknown vector parameter $\underline{\theta} = (\theta_1,\ldots,\theta_k)$. The marginal distribution of X_i depends on θ_i only , $i = 1,\ldots,k$. This paper deals with the problem of selecting the largest component of $\underline{\theta}$ and the analogous problem of selecting a subset of the components of $\underline{\theta}$ which includes the largest component. We consider the selection problem in a general decision theoretic framework and derive Bayes rules for selecting the largest component. The Bayes rules are shown to have certain optimal properties. The ordinary selection rules are shown to be Bayes rules, with respect to a special loss function.

*This work was supported by the U. S. Office of Naval Research under Contract N00014-75-C.

1. Introduction. Let $\underline{X} = (X_1,\ldots,X_k)$ be a random vector whose distribution depends on an unknown vector parameter $\underline{\theta} = (\theta_1,\ldots,\theta_k)$. The marginal distribution of X_i depends on θ_i only, for each $i = 1,\ldots,k$. We consider the problem of selecting the largest component of $\underline{\theta}$ given $\underline{x}$, an observed value of $\underline{X}$.

There are two formulations of the selection problem which have been generally considered in the literature. In one the goal is to select the largest component with a "high" probability. In the other the goal is to select a subset of the k components which includes the largest component with a high probability and includes any of the remaining components with a "low" probability. In the second case, the selection would be correct if the largest component is included in the selected subset.

In the standard formulation of the selection problem a minimum probability is pre-assigned, equal to P*, say, such that the probability of a correct selection (PCS) should be at least as large as P*. This is called the P*-condition. To meet the P*-condition it needs to find a "least favorable" configuration (lfc) of the parameter space for which the PCS is minimized. The lfc is found easily in some special cases which have been considered in the literature for the underlying distribution of $\underline{X}$. In other cases the minimization of the PCS is not so straightforward. Consider, for example, the case where $\underline{X}$ is distributed according to a multivariate normal distribution with mean $\underline{\theta}$ and covariance matrix Σ; where Σ is known. A simple rule can be given for selecting the largest component of $\underline{\theta}$ in the special case when the components of $\underline{X}$ have a common variance and are equi-correlated (see Gibbons, Olkin and Sobel (1977), § 15.2.1). It is not simple to find an optimal selection rule when Σ is defined more generally. The difficulty arises even in the case where the components of $\underline{X}$ are uncorrelated but they have unequal variances.

For this case various rules have been proposed in the literature for selecting a subset which includes the largest component of $\underline{\theta}$. Berger and Gupta (1980) have examined these rules and compared them, applying certain criterion of optimality.

In this paper we consider the selection problem in a Bayesian framework. The Bayes formulation involves the specification of a loss function and the assumption of a prior distribution for the parameter $\underline{\theta}$. Given the loss function and the prior distribution of $\underline{\theta}$, it is fairly easy to find an optimal selection rule. The optimal rule is called a Bayes rule. The Bayes solution does not involve the minimization problem of finding the least favorable configuration. Therefore, at least from the point of view of mathematical simplicity, a Bayes solution of the selection problem should be more attractive than the standard method, discussed above.

In the following section we give a decision theoretic formulation of the selection problem and derive the Bayes solution for a general loss function. We illustrate our result with an example from the multivariate normal distribution.

Berger and Gupta (1980) have considered a monotonicity property for an optimal selection rule. A rule is said to be just if it has that property. In Section 3 we show that our Bayes rules are just if certain conditions with regard to the distribution of $\underline{X}$ and the loss function are met.

2. Bayes selection rules. We formulate the selection problem in a decision theoretic framework, as follows: A rule for selecting the largest component of $\underline{\theta}$ (selecting a subset of the k components, which includes the largest component) is given by a vector $\phi(\underline{x}) = (\phi_1(\underline{x}),\ldots,\phi_k(\underline{x}))$, where $\phi_i(\underline{x})$ denotes the probability that the ith component is selected (included in the selected subset) when $\underline{x}$ is the observed value of $\underline{X}$. For the problem of selecting the largest component we have

$$(2.1) \qquad \sum_{i=1}^{k} \phi_i(\underline{x}) = 1 \quad , \quad \forall\, \underline{x} \quad .$$

First, consider the subset selection problem. We call it Problem I. Let $L_i(\underline{\theta})$ denote. The loss incurred due to including the ith component in the selected subset, and let $L_i^*(\underline{\theta})$ denote the loss due to cluding the ith component from the selected subset. The total loss due to selecting a subset (δ) is given by

$$(2.2) \qquad L(\delta,\underline{\theta}) = \sum_{i=1}^{k} \delta_i L_i(\underline{\theta}) + \sum_{i=1}^{k} (1-\delta_i) L_i^*(\underline{\theta})$$

$$= \sum_{i=1}^{k} \delta_i (L_i(\underline{\theta}) - L_i^*(\underline{\theta})) + \sum_{i=1}^{k} L_i^*(\underline{\theta})$$

where $\delta = (\delta_1,\ldots,\delta_k)$ and $\delta_i = 1(0)$ if the ith component is included in (excluded from) the selected subset. We assume that

$$(2.3) \qquad \sum_{i=1}^{k} (L_i(\underline{\theta}) - L_i^*(\underline{\theta})) \leq 0 \quad , \quad \forall\, \underline{\theta} \quad .$$

The above inequality implies that the loss due to including all the components in the selected subset is $\leq$ the loss due to excluding all the components from

the selected subset. Therefore, we include at least one component in the selected subset.

Next, consider the problem of selecting the largest component. We call it Problem II. Using the same generic notation for the loss function as in Problem I, we let $L_i(\underline{\theta})$ denote the loss due to selecting the ith component as the largest component, and let $L_i^*(\underline{\theta})$ be the loss due to not selecting the ith component. The total loss due to selecting a component (δ) is given by (2.2), where now $\delta_i = 1(0)$ if the ith component is selected (not selected) for the largest component.

Consider a special case of the loss function, given above. Let

$$(2.4) \qquad L_i(\underline{\theta}) = \begin{cases} 0 & \text{if } \theta_i = \theta_{[k]} \\ 1 & \text{if } \theta_i \neq \theta_{[k]} \end{cases}$$

$$L_i^*(\underline{\theta}) = c(1-L_i(\underline{\theta}))$$

where c is a positive number and $\theta_{[k]} = \max(\theta_1,\ldots,\theta_k)$. We let $c \geq k-1$ for Problem I and $c = 1$ for Problem II. In Problem I the value of c measures the loss due to excluding the largest component from the selected subset, relative to the loss due to including a wrong component in the selected subset. The inequality (2.3) holds since $c \geq k-1$.

The risk, that is, the expected loss due to a selection rule $\phi = \phi(\underline{x})$ is given by

$$(2.5) \qquad R_\phi(\underline{\theta}) = \sum_{i=1}^{k} (L_i(\underline{\theta}) - L_i^*(\underline{\theta}))\, E\, \phi_i(\underline{x}) + \sum_{i=1}^{k} L_i^*(\underline{\theta}) \quad .$$

In Problem I the risk for the loss function given by (2.4), is equal to the sum of c(1-PCS) and the expected number of wrong components included in the

selected subset. In Problem II the risk for the same loss function is equal to (1+c)(1-PCS).

Let $P_{\underline{\theta}}$ denote the conditional distribution of $\underline{X}$ given $\underline{\theta}$. We suppose that the distribution has a denisty $p_{\underline{\theta}}(\underline{x})$ with respect to a σ-finite measure μ on R^k. For the Bayes formulation of the selection problem we assume that $\underline{\theta}$ is distributed a' priori according to a probability distribution G, say. The optimal selection rule is a functional ϕ which minimizes the average of the risk function with respect to the given prior distribution of $\underline{\theta}$, given by

$$R_\phi = \int R_\phi(\underline{\theta})\, dG(\underline{\theta}).$$

Let

$$M_i(\underline{x}) = \int (L_i(\underline{\theta}) - L_i^*(\underline{\theta}))\, p_{\underline{\theta}}(\underline{x})\, dG(\underline{\theta}) \tag{2.6}$$

$$M(\underline{x}) = \min (M_1(\underline{x}), \ldots, M_k(\underline{x})) \quad . \tag{2.7}$$

By virtue of (2.5) a Bayes rule for Problem I is given by

$$\phi_i(\underline{x}) = \begin{cases} 1 & \text{if } M_i(\underline{x}) \leq 0 \\ 0 & \text{otherwise} \end{cases} \quad . \tag{2.8}$$

The Bayes rule for Problem II is given by

$$\phi_i(\underline{x}) = \begin{cases} 1 & \text{if } M_i(\underline{x}) = M(\underline{x}) \\ 0 & \text{otherwise} \end{cases} \quad . \tag{2.9}$$

If $M_1(\underline{x}) = M(\underline{x})$ for several values of i, we select the smallest among the tied values of i for the largest component. We note that the Bayes rules (2.8) and (2.9) are both non-randomized selection rules.

We illustrate our results by the following example.

Example. Let $\underline{X}$ be distributed according to a multivariate normal distribution $N(\underline{\theta},\Sigma)$, where the covariance Σ is a diagonal matrix, the ith element on the diagonal being denoted by σ_i^2. Let the loss function be given by (2.4) and let $\underline{\theta}$ be distributed a'priori according to $N(\underline{0}, \tau^2 I)$, where I denotes an identity matrix. Let $\phi(x)$ and $\Phi(x)$ denote the standard normal density and cdf, respectively, and let

$$\lambda_i^2 = \tau^2 / (\tau^2+\sigma_i^2) \quad .$$

By direct computation we get

$$(2.10) \quad M_i(\underline{x}) = p(\underline{x}) \left[1-(1+c) \int_{-\infty}^{\infty} \prod_{j=1(j\neq i)}^{k} \Phi \left(\frac{\lambda_i\sigma_i}{\lambda_j\sigma_j} u + \frac{\lambda_i^2 x_i - \lambda_j^2 x_j}{\lambda_j\sigma_j}\right)\phi(u)du\right]$$

where

$$p(\underline{x}) = \int p_{\underline{\theta}}(\underline{x}) \, dG(\underline{\theta})$$

$$= \left(\prod_{i=1}^{k} (2\pi(\sigma_i^2+\tau^2))^{-\frac{1}{2}} \right) \exp \left(- \frac{1}{2} \sum_{i=1}^{k} x_i^2 / (\sigma_i^2+\tau^2)\right)$$

denotes the marginal density of $\underline{X}$.

If we let $\tau \to \infty$, so that the prior distribution of $\underline{\theta}$ tends to be non-informative, then

$$(2.11) \quad M_i(\underline{x}) = p(\underline{x}) \left[1-(1+c) \int_{-\infty}^{\infty} \prod_{j=1(j\neq i)}^{k} \Phi \left(\frac{\sigma_i}{\sigma_j} u + \frac{x_i - x_j}{\sigma_j}\right)\phi(u) \, du\right] \quad .$$

Hence, the ith component is included in the selected subset if

$$(2.12) \quad \int_{-\infty}^{\infty} \prod_{j=1(j\neq i)}^{k} \Phi \left(\frac{\sigma_i}{\sigma_j} u + \frac{x_i - x_j}{\sigma_j}\right) \phi(u) \, du \geq \frac{1}{1+c} \quad .$$

The above inequality holds for sufficiently large values of $(x_i-x_j)/\sigma_j$, $j\neq i$. In Problem II we select the rth component if the quantity on the left side of (2.12) is maximized for i=r. For k=2 this quantity is equal to

$$\Phi\ (\ (x_i-x_j)\ /\ (\sigma_i^2+\sigma_j^2)^{\frac{1}{2}})\ .$$

Therefore, we select the component associated with the larger of the two values x_1 and x_2.

Let $\sigma_1 = \ldots = \sigma_k = \sigma$, say, giving $\lambda_1 = \ldots = \lambda_k = \lambda$, say. We have

$$(2.13)\quad M_i(\underline{x}) = p^0(\underline{x})\ [1-(1+c)\int_{-\infty}^{\infty} \prod_{j=1(j\neq i)}^{k} \Phi\ (u + \frac{\lambda}{\sigma}\ (x_i-x_j))\ \phi(u)\ du]$$

where $p^0(\underline{x})$ is obtained from $p\ (\underline{x})$ by substituting σ for σ_i , $i = 1,\ldots,k$. We find that $M_i(\underline{x}) \leq 0$ for $x_i-x_j \geq 0$, $j = 1,\ldots,i-1,\ i+1,\ldots,k$. Therefore, we include the ith component in the selected subset if none of these differences is negatively large. Also, $M_i(\underline{x})$ is minimized for the value of i associated with the largest component of $\underline{x}$. Therefore, in Problem II we select the component associated with the largest value among $x_1,\ldots,x_k$. These are ordinary selection rules. We see them as Bayes rules.

3. Just rules. First we define a stochastically increasing property (SIP) of a class of multivariate distributions. A set $A \subset R^k$ is said to be monotone if, if $\underline{x} \in A$ and $y_i \geq x_i$, $i=1,\ldots,k$ then $\underline{y} \in A$. Let $P_{\underline{\theta}}$ be a family of probability distributions on R^k indexed by a vector parameter $\underline{\theta} = (\theta_1,\ldots,\theta_k)$. Let $\Omega \subset R^k$ denote the parameter space. The family of distributions $P_{\underline{\theta}}$ is said to have SIP with respect to $\underline{\theta}$ if , if $\underline{\theta} \in \Omega$, $\underline{\theta}' \in \Omega$ and $\theta_i \leq \theta_i'$, $i = 1,\ldots, k$ then $P_{\underline{\theta}}(A) \leq P_{\underline{\theta}'}(A)$ for all monotone sets A. A character-

ization of the SIP, due to Lehmann (1955), is given as follows: A function $\psi(\underline{x})$ is said to be nondecreasing in $\underline{x}$ if, if $x_i \leq x_i'$, $i = 1,\ldots,k$ then $\psi(\underline{x}) \leq \psi(\underline{x}')$. A family of distributions $P_{\underline{\theta}}$ is said to have SIP with respect to $\underline{\theta}$ if and only if, if $\psi(\underline{x})$ is nondecreasing in $\underline{x}$ then $\int \psi(\underline{x})\, d\, P_{\underline{\theta}}(\underline{x})$ is nondecreasing in $\underline{\theta}$.

Now we define a just rule. A selection rule ϕ is said to be just if $\phi_i(\underline{x})$ is nondecreasing in x_i and nonincreasing in x_j $(j \neq i)$ for $i, j = 1,\ldots,k$. Theorem 1 below, shows that the Bayes rules given by (2.8) and (2.9) are just if the following assumptions are valid.

Assumption 1 - The posterior distribution of $\underline{\theta}$ given $\underline{x}$, has SIP with respect to $\underline{x}$.

Assumption 2 - The function $L_i(\underline{\theta}) - L_i^*(\underline{\theta})$ is nonincreasing in θ_i and nondecreasing in $\theta_i (j \neq i)$.

Assumption 3 - $M_i(\underline{x})$ is a continuous function of $\underline{x}$.

Assumption 3 is valid if, for example, the loss functions $L_i(\underline{\theta})$ and $L_i^*(\underline{\theta})$ are bounded and $p_{\underline{\theta}}(\underline{x})$ is continuous in $\underline{x}$ uniformly for $\underline{\theta} \in \Omega$.

Theorem 1. If assumptions 1 and 2 hold then the Bayes rule (2.8) is just. If moreover Assumption 3 holds then the Bayes rule (2.9) is just.

Proof: From the characterization of the SIP given above, and Assumptions 1 and 2, it follows that $M_i(\underline{x})$ as given by (2.6), is nonincreasing in x_i and nondecreasing in $x_j (j \neq i)$. Therefore, the function $\phi_i(\underline{x})$ as given by (2.8), is nondecreasing in x_i and nonincreasing in $x_j (j \neq i)$. Hence, the Bayes rule (2.8) is just if Assumptions 1 and 2 hold. If moreover, Assumption 3 holds then it follows from the continuity and monotonicity property of $M_i(\underline{x})$ that $\phi_i(\underline{x})$, given by (2.9), is nondecreasing in x_i and nonincreasing in $x_j (j \neq i)$. Hence, the Bayes rule (2.9) is just. □

In the application of Theorem 1 it would be interesting to find for a given family of distributions $P_{\underline{\theta}}$ which is stochastically increasing in $\underline{\theta}$, the family of prior distributions for which Assumption 1 holds. We have not investigated this problem to any length. We discuss below some cases in which an appropriate prior distribution can be found for which Assumption 1 holds.

Clearly, $P_{\underline{\theta}}(\underline{x})$ has SIP with respect to $\underline{\theta}$ if $\underline{\theta}$ is a location parameter of the conditional distribution. The posterior distribution of $\underline{\theta}$ with respect to a non-informative prior distribution G which is uniform on R^k, has SIP with respect to $\underline{x}$. Similarly, $P_{\underline{\theta}}(\underline{x})$ has SIP with respect to $\underline{\theta}$ if $\underline{\theta}$ is a scale parameter (component-wise) of the conditional distribution. The posterior distribution of $\underline{\theta}$ with respect to the non-informative prior on R^k with density function $g(\underline{\theta}) \propto (\theta_1 \dots \theta_k)$ has SIP with respect to $\underline{x}$.

The prior distributions for the location and scale parameters considered above, which lead to the posterior distribution with SIP are both improper distributions. We give now an example of a proper prior distribution for a location parameter. Let $P_{\underline{\theta}}$ denote the multivariate normal distribution $N(\underline{\theta},\Sigma)$ and let $\underline{\theta}$ be distributed a priori according to $N(\underline{0},\Omega)$. A' posteriori, $\underline{\theta}$ is distributed according to the normal distribution $N(A\,\underline{x}\,,\,(\Sigma^{-1}+\Omega^{-1})^{-1})$, where

$$A = (\Sigma^{-1} + \Omega^{-1})^{-1}\,\Sigma^{-1}.$$

If the elements of A are non-negative then the posterior distribution of $\underline{\theta}$ has SIP with respect to $\underline{x}$. The elements of A are positive if the covariance matrices $\Sigma = (\sigma_{ij})$ and $\Omega = (\omega_{ij})$ are given by $\sigma_{ii} = 1$, $\sigma_{ij} = \rho$ $(i \neq j)$, $\rho < 0$, $\omega_{ii} = \frac{1}{\lambda}$ and $\omega_{ij} = 0$ $(i \neq j)$. The matrix A is given by

$$A = (1+\lambda(1-\rho)^{-1})^{-1}\,\left(I - \frac{\lambda\rho}{1+\lambda+\lambda\rho(k-1)}\,J\right)$$

where I denotes the identity matrix and J denotes a matrix whose elements are each equal to 1.

References

[1] Berger, R. L. and Gupta, S. S. (1980). Minimax subset selection with applications to unequal variances (unequal sample size) problems. Scand. J. Statist. 7: 21-26.

[2] Gibbons, J. D., Olkin, I. and Sobel, M. (1977). Selecting and Ordering Populations - A New Statistical Methodology. John Wiley and Sons, New York.

[3] Lehmann, E. L. (1955). Ordered families of distributions. Ann. Math. Statist. (26) 399-419.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: N127	2. GOVT ACCESSION NO.: AD. A112 976	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): A BAYES RULE FOR SELECTING THE LARGEST COMPONENT.		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER: TR #365
7. AUTHOR(s): Khursheed Alam		8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0451
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Clemson University, Dept. of Mathematical Sciences, Clemson, South Carolina 29631		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 365-049
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Code 434, Arlington, Va. 22217		12. REPORT DATE: 16 July 1981
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report): Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Ranking and Selection, Least favorable configuration, Bayes rules.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let $\underline{X} = (X_1,\ldots,X_k)$ be a random vector whose distribution depends on a vector parameter $\underline{\theta} = (\theta_1,\ldots,\theta_k)$. The marginal distribution of X_i depends on θ_i. Bayes rules are given for selecting the largest component of $\underline{\theta}$ and for selecting a subset of the components of $\underline{\theta}$, including the largest component. The given rules are shown to have certain optimal properties.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)